Chiara Pederzoli (Italy), Costanza Torricelli (Italy)

A parsimonious default prediction model for Italian SMEs

Abstract

In the light of the fundamental role played by small and medium enterprises (SMEs) in the economy of many countries, including Italy, and of the specific treatment of this issue within the Basel II regulation, the aim of this paper is to build a default prediction model for the Italian SMEs. Specifically, this study develop a logit model based on financial ratios. Using the AIDA database, the authors focus the attention on a specific region in Italy, Emilia Romagna, where SMEs represent the majority of firms. The paper finds that a parsimonious model, based on only four explanatory variables, fits well the default data and provides results consistent with structural models of the Merton type.

Keywords: probability of default (PD), SME, Basel II.
JEL Classification: G24, G32, C25.

Introduction

Small and medium enterprises (SMEs) play a very important role in the economic system of many countries and particularly in Italy. One of the main problems of Italian SMEs is to recover money to finance their investments. The role of banks in Italy is very important, since they are the only subject issuing loans directly to SMEs and to this end they need models for the estimation of the probability of default (PD). An additional reason to develop specific models for SMEs lies in the Basel II regulation, since the estimation of the obligors' PD is a fundamental issue for banks adopting the internal ratings-based (IRB) approach. Basel II, in fact, requires these banks to set up a rating system and provides a formula for the calculation of minimum capital requirements, where the PD is the main input. Moreover, the regulation recognizes a different treatment for the exposures towards SMEs, which benefit from a reduction of the capital requirement proportional to their size.

Based on the above premises, the aim of this work is to develop a default prediction model for the Italian SMEs, focusing the attention on a specific geographic area, namely the Emilia Romagna region, where SMEs represent the firms' majority.

The model we propose is a logit model based on balance-sheet data. A wide range of models for the estimation of the corporates' default probability have been developed. These models can be classified according to the type of data required. The models for pricing risky debt, having their milestone in the Merton model, are based on market data and, therefore, they are not suitable for small (not quoted) enterprises. On the contrary, statistical models, such as those based on discriminant analysis and binary choice models, mainly use accounting data which are available for all enterprises regardless of their size. This paper focuses on balance sheet data which are public so that the model proposed lends itself to be used not only by banks but by any economic agent who may be interested in the firm's credit quality.

The paper is organized as follows. The literature related to default prediction, in particular for SMEs, is briefly presented in Section 1. Section 2 illustrates relevant issues related with the dataset used and the approach adopted, while Section 3 presents the results obtained. The last Section concludes.

1. Literature overview

There is a wide range of default prediction models, i.e. models that assign a probability of failure or a credit score to firms over a given time horizon. The literature on this topic has developed especially in connection with Basel II, which allows banks to set up an internal rating system, that is a system to assign ratings to the obligors and to quantify the associated PDs. As stressed in the introduction, some sophisticated models available in the literature can be used only if market data on stocks (structural models) or corporate bonds and asset swaps (reduced-form models) are available. As for SMEs, for which market data are generally not available, either heuristic (e.g., neural network) or statistical models can be applied.

Beaver (1966) and Altman (1968) first used discriminant analysis (DA) to predict default. In order to overcome the limits inherent in DA (e.g., strong hypotheses on explanatory variables, equal variance-covariance matrix for failed and not failed firms), logit and probit models have been widely adopted[1]. An important advantage of the latter models is the immediate interpretation of the output as a default probability. A seminal paper in this respect

The authors gratefully acknowledge financial support from MIUR-PRIN 2007. We wish to thank Andrea Mazzali and Maria Teresa Palumbo for valuable research assistance and conference participants of XXXIII AMASES Conference (Parma) for helpful comments and suggestions. Usual caveat apply.

[1] A number of papers, among which Lennox (1999) and Altman and Sabato (2007), show that probit/logit models outperform DA model in default prediction.

is the one by Ohlson (1980), who analyzed a dataset of U.S. firms over the years of 1970-1976 and estimated a logit model with nine financial ratios as regressors. Despite the diffusion of the pricing models based on market data, the logit/probit models, based on accounting data, are nowadays widely used. Recently Beaver (2005), by analyzing a dataset of U.S. firms over the period of 1962-2002, has shown that balance sheet financial ratios still preserve their predictive ability, even if market-based variables partly encompass accounting data.

A relatively new approach, based on machine learning, is the maximum expected utility (MEU). This model, developed at the Standard & Poor's Risk Solutions Group (Friedman and Sandow, 2003), is based on the maximization of the expected utility of an investor who chooses her investment strategy based on her beliefs and on the data. Marassi and Pediroda (2008) applies this approach to a dataset of Italian firms.

Focusing on SMEs, a few recent works use logit/probit models, or some evolution of the same, for PD estimation: Altman & Sabato (2007) use a dataset of U.S. SMEs; Altman and Sabato (2005) analyze separately U.S., Australian and Italian SMEs; Behr and Güttler (2007) and Fantazzini and Figini (2009) analyze German data; Fidrmuc and Heinz (2009) use data from Slovakia. Despite some differences among these analyses, a convergence emerges on some types of financial indicators, which can be grouped into five categories: leverage, liquidity, profitability, coverage, activity (Altman and Sabato, 2007).

2. The construction of the data set

The sample for the empirical analysis is entirely drawn from AIDA, a financial database powered by Bureau Van Dijk which contains the balance sheet data of all the Italian firms. Indeed, we use public data only, while banks usually build their models on private data (e.g., default on single bank loans) taken from credit registers.

Given the aim of our research, we restrict our attention to SMEs. In order to construct an appropriate data set, there are a number of issues we have to tackle. The first one is the very same definition of SME, for which we stick to the Basel II rule. The definition given by the European Union[1] refers both to the number of employees and to the sales: firms are considered small, if they have less than 50 million euros in sales or less than 250 employees. The Basel Committee on Banking Supervision (BCBS), for the purpose of capital requirements, imposes a criterion based on sales only to discriminate between SMEs and corporates: firms with annual sales less than 50 million euros are considered SMEs and this imply for the intermediary a reduction in capital requirement[2] proportional to the firm's size. In our sample we have included only firms with annual sales lower than 50 million euros[3], consistently with the Basel II definition. This choice is motivated by the ultimate aim of this work: the estimated PDs are used in fact as input in the Basel II capital requirement formula.

As for the geographic focus, we concentrate on a particular area, the Emilia Romagna region, in order to develop a model able to capture the specific features of the firms in this region, since it is highly representative of SMEs.

In our sample we consider balance sheet data for 2004 to estimate the one-year PD. Another relevant issue is the definition of default to be used in the classification. In order to classify defaulted firms in our sample, we need, first of all, to adopt a definition of default, since the literature does not provide a univocal one. We refer to Altman and Hotchkiss (2006) for the various definition: *failure, insolvency, default* and *bankruptcy*, which are used interchangeably in the literature but have different meaning and refer to different situations in different countries' bankruptcy law.

The BCBS (2006) adopts a wide default definition in that "a default is considered to have occurred with regard to a particular obligor when either or both of the two following events have taken place:

- the bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realising security (if held);
- the obligor is past due more than 90 days on any material credit obligation to the banking group. Overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstandings.

Often default definitions for credit risk models concern single loan defaults of a company versus a bank, as also emerges from the above Basel II instructions. This is the case for banks building models based on their portfolio data, that is relying on

[1] Commission recommendation 96/280/EC of April 3, 1996, updated in 2003/361/EC of May 6, 2003. See http://europa.eu/scadplus/leg/en/lvb/-n26026.htm.

[2] The reduction applies to the capital function through the correlation, which is reduced by a maximum of 0.04 for the smallest firms. This correction is justified by the assumption that defaults of small firms are less correlated and, therefore, less risky on the whole for the portfolio.

[3] From the SMEs original data set we deleted firms with sales less than 100 000 euros since we believe that such small firms may be very different from typical firms working in industrial sectors in terms of operational, financial and economic features.

single loans data which are reserved (e.g., Altman and Sabato (2005) develop a logit model for Italian SMEs based on the portfolio of a large Italian bank). However, traditional structural models (i.e. Merton type models) refer to a firm-based definition of default: a firm defaults when the value of the assets is lower than the value of the liabilities, that is when equity is negative.

In this work default is intended as the end of the firm's activity, i.e. the status, where the firm needs to liquidate its assets for the benefit of its creditors. In practice, we consider a default occurred when a specific firm enters a bankruptcy procedure as defined by the Italian law. The reason for this choice lies in the data availability but it is also motivated by the objective of the paper: our aim is to define a model, based on public and accessible data, that measures the health state of the firms and enables any economic subject interested in a specific firm's health (i.e. suppliers, customers, lenders, etc.) to estimate the probability of a particular firm to get bankrupted.

In practice, in order to create our sample from the AIDA database, we associate the event of default to the absence of deposited balance sheet[1]: for the Italian law, firms must not deposit their balance sheet at the firms registry (*Registro delle Imprese*[2]) if, in a particular year, a bankruptcy proceeding starts. In general, a bankruptcy proceeding occurs when a firm is configured as an insolvent debtor and it can start after a specific request of the insolvent debtor, one or more creditors, the Public Prosecutor or the Law Court. According to these observations, we build our sample for the year 2004 by focusing on two groups of firms:

- Active firms: firms that are currently operative (i.e. not bankrupted)[3].
- Bankrupted firms: firms that are currently failed and whose last balance sheet was registered in 2005.

We assume that failed firms which deposited their last balance sheet in 2005 entered the bankruptcy proceeding in 2006. Therefore, we analyze the balance sheet data from one to two years before bankruptcy to estimate the probability of default.

The total default rate in the sample is about 0.6 %[4].

[1] Even if AIDA provides a flag to distinguish currently failed firms, it is not possible to select firms failed in a particular year automatically.

[2] The "Registro delle Imprese" is the Italian registry office which collects the balance sheet information of all the Italian firms.

[3] The current status refers to the time of the data collection, i.e. January 2008.

[4] It has to be noted that the default rate is very low if compared with some other works: this difference is due to the definition of default adopted, which is a consequence of the type of data available. For example, in Altman and Sabato (2005) any delay (more than 90 days) in the payments is counted as default, while in the present paper only the firms actual defaults are considered.

3. The empirical analysis.

In line with most of the literature based on accounting data, we use a binary logistic regression model. The default probability in a logit model is estimated by equation (1):

$$PD_i = P(Y_{i,t+1} = 1) = \frac{\exp(\alpha + \sum_{k=1}^{R} \beta_k X_{ik,t})}{1 + \exp(\alpha + \sum_{k=1}^{R} \beta_k X_{ik,t})}, \quad (1)$$

where:

$$Y_{i,t+1} \quad {}_{i=1,\dots,n} = \begin{cases} 1 \text{ if obligor } i \text{ defaults in } t+1, \\ 0 \text{ if obligor } i \text{ does not defaults in } t+1, \end{cases}$$

$$X_{ik,t} \quad {}_{i=1,\dots,n} = k^{th} \text{ regressor for obligor } i \text{ in } t.$$

We quantify the dependent variable according to the definition of default given in Section 2, while we consider balance sheet variables as regressors. The main issue is precisely the selection of appropriate and informative balance sheet variables, as explained in the following subsection.

3.1. Selection of the predictors. In order to select the appropriate regressors, we start by considering a number of variables which have been largely used in the default prediction literature, namely we choose 16 financial ratios, presented in Table 1, related to the main aspects of a company's financial profile (leverage, liquidity, profitability, coverage, activity).

Table 1. List of candidate predictors

Financial ratio	Categoria
Inventory/sales (*IS*)	ACTIVITY
Sales/asset (*SALESA*)	ACTIVITY
Short term debt/equity (*STDE*)	LEVERAGE
Long term liabilities/asset (*LTLA*)	LEVERAGE
Equiy/asset (*EQUITYA*)	LEVERAGE
Ebit/asset (*EBITA*)	PROFITABILITY
Ebit/sales (*ES*)	PROFITABILITY
Economic value addded/asset (*EVAA*)	PROFITABILITY
Net income/asset (*NIA*)	PROFITABILITY
Working capital/asset (*WCA*)	LIQUIDITY
Cash/asset (*CA*)	LIQUIDITY
Working capital/sales (*WCA*)	LIQUIDITY
Working capital/current liabilities (*WCC*)	LIQUIDITY
Cash/current/liabilities (*CCL*)	LIQUIDITY
Current liabilities/asset (*CLA*)	LIQUIDITY
Ebit/interest expenses (*EIE*)	COVERAGE

We select among these candidate predictors by means of a backward elimination procedure based on the Schwartz information criterion (SIC). The resulting model is illustrated in Table 2. The estimation results show that all the coefficients display the expected sign and are significant.

The equity ratio (*EQUITYA*) indicates the relative proportion of equity used to finance the company's assets. In general, we expect that a higher equity ratio implies a decrease in an SME's default risk and the model confirms this presumption. The current ratio measures whether a firm has enough resources to pay its debts over the next 12 months. The ebit/asset ratio measures the ability of generating income without tax distortion: the higher this ratio, the more healthy should the firm be and, hence, the lower is the PD. The long-term liabilities to asset ratio quantifies the long term debt compared to the short term one: higher long-term liabilities means (by construction) lower short-term ones, and, for this reason, the higher is this ratio the lower is the PD. A high value for the sales/asset indicator means good performances on the market and, therefore, a low PD.

Table 2. Estimation output

Estimated equation:

$$PD = 1/(1+\exp(2.86+3.46\ LTLA+ +3.52\ EBITA+11.18\ EQUITYA\ +0.43\ SALESA))$$

Variable	Estimated coefficient	Std.error (Huber /White)	Z-stat.	Prob.
CONSTANT	-2.8654	0.3467	-8.2679	0.000
EQUITYA	-11.1832	2.9199	-3.8299	0.000
EBITA	-3.5190	1.3478	-2.6110	0.009
LTLA	-3.4596	0.7688	-4.4999	0.000
SALESA	-0.4315	0.2393	-1.8034	0.071
Mean dep. var.	0.00573	S.D. dep. var.	0.07547	
S.E. regession	0.07201	Akaike I. C.	0.05913	
Sum sq. res.	85.9835	Schwarz I.C.	0.06146	
Log likelihood	-485.410	Hannan Quinn I.C.	0.05990	
Restr. log lik.	-585.159	Avg. log lik.	-0.02927	
LR stat. (5 d.f.)	199.498	Mc Fadden R-sq	0.1705	
Prob. (LR stat.)	0.000			

3.2. Model performance. The performances of the default prediction model can be measured in different ways: an exhaustive presentation of the available validation techniques can be found in BCBS (2005).

Consistently with most of the literature, we evaluate the performance of our model by means of the cumulative accuracy profile (CAP) and the associate accuracy ratio (AR), which measures the ability of the model to maximize the distance between the defaulted and non-defaulted firms[1]. Figure 1 shows the in sample CAP curve for our model; the associate AR is 66.84%.

[1] See Sobehart et al. (2001) and Engelman et al. (2003) for a discussion of the CAP curve and the accuracy ratio.

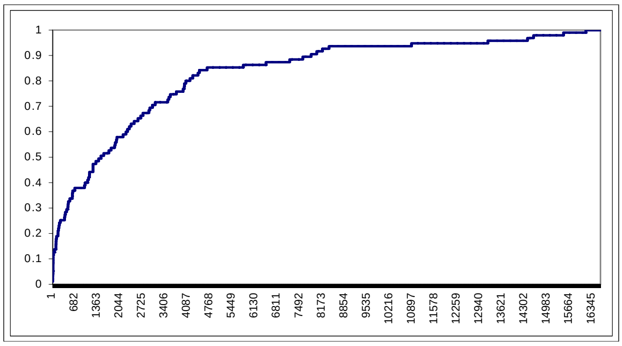

Fig.1. Cumulative accuracy profile of the model

While common goodness of fit measures for binary choice models rely on the choice of a particular cut-off value to discriminate between the two states, the AR indicator is free of arbitrary choices. Table 3 shows the error rates for some values of the discriminating cut-off: obviously type 1 error increases with increasing cut-off values, while type 2 error decreases; the average error rate is low when the cut-off value is fixed at the level of the sample default rate.

Table 3. Error rates

Cut-off	Type 1 error rate	Type 2 error rate	Avg error rate
0.006	14.74%	30.82%	22.78%
0.01	31.58%	17.37%	24.47%
0.05	87.37%	0.1%	43.73%
0.1	87.37%	0.03%	43.70%

Note: Type 1 error refers to failed firms classified as not failed; type 2 error refers to not failed classified as failed.

Conclusions

Two objects are the fundamental premises for the analyses presented in this paper. First, small and medium enterprises which are the backbone of the Italian economy – particularly in some regions such as Emilia Romagna – rest predominantly on the banking sector for their funding needs. Second, the peculiarity of SMEs in terms of credit assessment is highlighted by their specific treatment within the Basel II regulation for minimum capital requirements. These two premises call for the need to reconsider PD estimation models, which, in the absence of market data, have to rely on balance sheet data.

To this end, we have developed a logit default prediction model for the Italian SMEs in the Emilia Romagna region based on publicly available balance sheet data. The results obtained show that the model behaves fairly well in sample and, thus, confirm the validity of limited dependent variable models with financial ratios as predictors to represent default events. We find that a parsimonious model with four predictors, namely the equity ratio, the long term liabilities over asset ratio, the ebit over asset ratio and the sales over asset ratio, is sufficient to fit default events in our sample. In particular, the equity ratio on its own explain very well defaults: this means that the idea underlying the Merton approach, based on the relation between assets, liabilities and equity, holds also for SMEs. Thus, even if the application of the Merton model is generally prohibited for SMEs since it requires market data, our results show some consistency between reduced form and structural models.

References

1. Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, Vol. 13, pp. 589-609.
2. Altman, E.I., Sabato, G. (2005). Effects of the new Basel capital accord on bank capital requirements for SMEs, *Journal of Financial Services Research*, Vol. 28, pp. 15-42.
3. Altman, E.I., Hotchkiss, E. (2006). Corporate financial distress and bankruptcy: predict and avoid bankruptcy, analyze and invest in distressed debt, 3rd Edition, Hardcover: Wiley Finance.
4. Altman, E.I., Sabato, G. (2007). Modeling credit risk for SMEs: evidence from the U.S. market, *Abacus*, Vol. 43 (3), pp. 332-357.
5. Basel Committee on Banking Supervision (BCBS) (2005). Studies on the validation of internal rating systems, Working Paper N. 14, February.
6. Basel Committee on Banking Supervision (2006). International convergence of capital measurement and capital standards: a revised framework – comprehensive version, Bank for International Settlements, June.
7. Beaver V. (1966). Financial ratios as predictors of failure, *Journal of Accounting Research*, Vol. 5, pp. 71-111.
8. Beaver V. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy, *Review of Accounting Studies*, Vol. 10, pp. 93-122.
9. Behr, P Güttler, A. (2007). Credit risk assessment and relationship lendig: an empirical analysis of German small and medium-sized enterprises, *Journal of Small Business Management*, Vol. 45(2), pp. 194-213.
10. Engelmann, B. Hayden, E. Tasche, D. (2003). Measuring the discriminative power of rating systems, Deutsche Bundesbank, Discussion paper Series 2: Banking and Financial Supervision No 01/2003.
11. Fantazzini D., Figini S. (2009). Default forecasting for small-medium enterprises: does heterogeneity matters? *International Journal of Risk Assessment and Management*, Vol. 11 (1-2), pp. 38-49.
12. Fidrmuc J., Heinz C. (2009). Default rates in the loan market for SMEs: evidence from Slovakia, University of Munich, IFO Working Paper, №. 72.
13. Friedman C., Sandow S. (2003). Learning probabilistic models: an expected utility maximization approach, *Journal of Machine Learning Research*, Vol. 4, pp. 257-291.
14. Lennox, C. (1999). Identifying failing companies: a reevaluation of the logit, probit and DA approaches, *Journal of Economics and Business*, Vol. 51, pp. 347-364.
15. Marassi, D. Pediroda, V. (2008). Risk insolvency predictive model maximum expected utility, *International Journal of Business Performance Management*, Vol. 10 (2/3), pp. 174-190.
16. Ohlson J. (1980). Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, Vol. 18, pp. 109-131.
17. Sobehart, J. Keenan, S. Stein, R. (2001). Benchmarking quantitative default risk models: a validation methodology, Algo Research Quarterly, Vol. 4, Nos.1/2 March/June 2001, pp. 57-72.